Multiplicity of solutions to GW-type approximations

F. Tandetzky, J. K. Dewhurst, S. Sharma and E. K. U. Gross Max-Planck-Institut für Mikrostrukturphysik, Weinberg 2, D-06120 Halle, Germany (Dated: March 4, 2013)

We show that the equations underlying the GW approximation have a large number of solutions. This raises the question: which is the physical solution? We provide two theorems which explain why the methods currently in use do, in fact, find the correct solution. These theorems are general enough to cover a large class of similar algorithms. An efficient algorithm for including self-consistent vertex corrections well beyond GW is also described and further used in numerical validation of the two theorems.

INTRODUCTION

The GW approximation[1] is a many-body technique used typically for the calculation of spectral density functions of solids. Its accuracy for the determination of band gaps of insulators as well as its parameter-free nature makes it a very attractive method in condensed matter physics. The self-consistent GW approximation is a fixed point method involving multidimensional objects like the Green's function and the self-energy, and it was recently demonstrated for an artificial one-point-model that this fixed point is not unique and that a different way of iterating the equations leads to a different solution[2]. It was further argued that including vertex corrections could exacerbate this non-uniqueness problem and lead to several solutions. Given that the GW approximation is a state of the art method for band structure calculations, such an ambiguity is a serious issue. It is also important to understand how one can go beyond GW without running into unphysical solutions.

In this article we describe a new algorithm for computing the self-energy well beyond the GW approximation. We investigate the nature of the additional solutions numerically and further provide general theorems, that explain why the methods currently in use to solve the GW equations do indeed lead to a unique solution. Since we cover a large class of approximations these results not only validate the GW calculations that are done, but they also provide conditions on approximations going beyond GW for obtaining a meaningful result.

The starting point are the Hedin equations, which appear as Eqs. (A22)-(A25) in the appendix of the 1965 article of Hedin[3]. We rewrite them here in modern notation:

$$\Gamma(1,2;3) = \Gamma_0(1,2;3) + \frac{\delta \Sigma(1,2)}{\delta V(3)}$$
 (1)

$$\Sigma(1,2) = i\lambda \int G(1,4)W(1^+,3)\Gamma(4,2;3)d(3)d(4) \quad (2)$$

$$\Pi(1,2) = -i\lambda \int G(2,3)G(4,2^+)\Gamma(3,4;1)d(3)d(4) (3)$$

$$\frac{\delta G(1,2)}{\delta V(3)} = \int G(1,4)G(5,2)\Gamma(4,5;3)d(4)d(5) \tag{4}$$

$$\frac{\delta W(1,2)}{\delta V(3)} = \int W(1,4)W(5,2)\frac{\delta \Pi(4,5)}{\delta V(3)}d(4)d(5). \tag{5}$$

In these equations G is the Green's function, W is the renormalized Coulomb propagator [4], Σ is the selfenergy, Π is the polarization, $\Gamma_0(1,2;3) = \lambda \delta(1,2)\delta(1,3)$ is the bare vertex, Γ is the renormalized vertex, λ is the coupling constant, and the potential differential δV is the sum of the external and Hartree contributions. The notation (1) $\equiv x_1 \equiv (\mathbf{r}_1, \sigma_1, t_1)$ is used throughout [5]. Note that we introduced the coupling constant λ such that in the non-interacting case the vertex function vanishes [5]. Instead one could also define it in such a way that the Coulomb propagator vanishes. Using simple transformations (introduced later in Eq. (12); use a = 1, b = λ^2 , $c = \lambda^{-1}$) one can show that these definitions are in fact equivalent. The physical meaning of this is, that it does not matter whether we define the non-interacting limit by switching off the interaction of the electrons with the photons or by setting the photon propagator to zero.

Hedin has shown, how one can gain an expansion of the vertex and hence of Σ and Π in terms of the renormalized quantities G and W using these equations. That way one gets Σ and Π as functionals $\Sigma[G,W]$ and $\Pi[G,W]$. In addition to the five Hedin equations are the two coupled Dyson equations:

$$G(1,2) = G_0(1,2) + \int G_0(1,3)\Sigma(3,4)G(4,2)d(3)d(4)$$

$$W(1,2) = W_0(1,2) + \int W_0(1,3)\Pi(3,4)W(4,2)d(3)d(4),$$
(6)

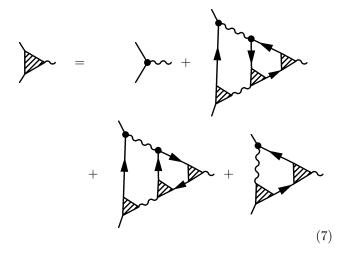
where G_0 is the Green's function of the non-interacting system (which includes the Hartree potential) and $W_0(1,2) = \frac{\delta(t_1-t_2)}{|\mathbf{r}_1-\mathbf{r}_2|}$ is the bare Coulomb propagator. Solving the Dyson equations in conjunction with the Hedin equations yields the functionals $G[G_0, W_0]$ and $W[G_0, W_0]$.

The separation of the problem into equations for Σ and Π as functionals of G and W, as well as G and W as functionals of G_0 and W_0 is an important conceptual step. In a later article by Hedin and Lundqvist[6], the equations are combined and the functional derivative $\delta\Sigma/\delta G$ is introduced. We would like to stress that this vertex equa-

tion together with Eqs. (2), (3) and the Dyson equations are not immediately useful. One also needs the equations (4) and (5) in order to get an expansion of the vertex beyond GW.

ALGORITHMS FOR HEDIN'S EQUATIONS

Almost all practical calculations of Hedin's equations use the GW approximation. This amounts to approximating the full vertex Γ by the bare vertex, and thus the self-energy takes on the simple form $\Sigma(1,2) =$ $i\lambda^2 G(1,2)W(1,2)$. In the present work we describe a new algorithm for solving Hedin's equations which includes corrections far beyond the GW approximation. From the outset we will insist that the computational storage requirements scale as N^3 and the number of operations scale as N^4 . These conditions are met by Algorithm 1 (see structogram) for calculating $\Sigma[G, W]$ and $\Pi[G, W]$ using a non-trivial vertex. Note that all the relations in the algorithm are exact apart from the two which have the derivatives of Γ removed – these would require N^4 storage. Solving Algorithm 1 together with the Dyson equations we refer to as the 'Starfish' algorithm. It is straight-forward to work out which diagrams this algorithm corresponds to: finding the self-consistent solution to Starfish is equivalent to solving



for the vertex, using this to find Σ and Π and subsequently G and W. The entire procedure is performed self-consistently.

NUMBER AND STABILITY OF SOLUTIONS

Discretization of space-time is required for the purpose of closely examining solutions to these equations. There is some ambiguity as to how this should be done, but at this level, we merely assume that there are N space-time points in total, and that limits in the time variable such

Algorithm 1 Hedin equations solver for $\Sigma[G,W]$ and $\Pi[G,W]$. Here the shorthand $G'(1,2;3) = \delta G(12)/\delta V(3)$, etc., is used and the function arguments are omitted.

Require:
$$G$$
 and W
Set $\Gamma = \Gamma_0$
repeat
$$G' = \int GG\Gamma$$

$$\Pi' = -i\lambda \int \left(G' G\Gamma + GG' \Gamma + G/G \Gamma' \right)$$

$$W' = \int W W \Pi'$$

$$\Sigma' = i\lambda \int \left(W' G\Gamma + W G' \Gamma + W/G \Gamma' \right)$$

$$\Gamma = \Gamma_0 + \Sigma'$$
until Γ converged
$$\Sigma = i\lambda \int W G\Gamma$$

$$\Pi = -i\lambda \int G G\Gamma$$

as $G(4,2^+)$ are taken to mean G(4,2). Also the Dirac delta function in Eq. (1) now becomes a Kronenker delta function. In doing so one loses the physical meaning of these equations, but we will assume that the true physical solution can be recovered in the continuum limit when $N \to \infty$.

Irrespective of whether the *GW* approximation, Starfish or some other truncation is used, the equations to be solved form a closed system of *polynomial equations* enabling us to prove general theorems about such a system.

To do so we define a concise notation. Let

$$F \equiv (G, W, \Pi, \Sigma, \Gamma) \in \mathbb{C}^n,$$

be the vector of all dependent (or unknown) variables. Here $n = 4N^2 + N^3$ is the number of unknowns. We regard the tensors G_0 and W_0 (and the bare vertex) as known and fixed. Equations (2), (3), (6) and a vertex equation like Eq. (7) can now be written in a compact form as $F = \mathfrak{g}(F)$ or $\mathfrak{h}(F) = 0$ with $\mathfrak{h}(F) \equiv \mathfrak{g}(F) - F$, where $\mathfrak{g} = (\mathfrak{g}_1, \dots, \mathfrak{g}_n)$ is a set of polynomials in several variables.

It is important to point out that most of the following considerations do not depend on the precise form of the vertex equation. For example the trivial vertex equation $\Gamma = \Gamma_0$ corresponding to the GW approximation is also allowed here. In this case one can also eliminate the vertex from the equations and redefine F, \mathfrak{g} and \mathfrak{h} in order to include only the smaller set of quantities and equations. Similarly, if one were to fix W in the Starfish algorithm then F would be (G, Σ, Γ) and the equations for W and Π could be eliminated.

Since we are interested in the dependence of these equations and their solutions on the coupling strength λ , the equations to be solved are

$$F = \mathfrak{g}^{\lambda}(F)$$
 or $\mathfrak{h}^{\lambda}(F) = 0.$ (8)

Multiple fixed points

We now want to determine the number of solutions to Eq. (8). An upper bound is provided by Bézout's theorem[7] which states that the maximum number of solutions to a system of polynomial equations (if finite) is equal to the product of the total degree of each equation. Thus the GW approximation with fixed W has at most 2^{N^2} solutions, and the Starfish algorithm, also for fixed W, has at most $7^{N^3}2^{2N^2}$ solutions. The Bézout bound fails to take into account sparsity in the polynomial equations and therefore also degeneracy of the roots. Buchberger's algorithm is a method of systematically determining the exact number of roots by decomposing the equations into a Gröbner basis[8]. This procedure is, however, computationally very demanding and can be performed only for small (and therefore nonphysical) N. For example, when N=2 the GW approximation with fixed W has precisely 6 solutions. Likewise, for N=1 the Starfish algorithm yields 3 solutions. From these considerations, it seems quite surprising that selfconsistent GW works at all for realistic values of N. We will now provide two theorems that may explain this apparent success.

Theorem 1. Let $\mathfrak{h}^{\lambda} = (\mathfrak{h}_1^{\lambda}, \dots, \mathfrak{h}_n^{\lambda}), \ \lambda \in [0, 1]$ be a set of polynomials $\mathfrak{h}_k^{\lambda} : \mathbb{C}^n \to \mathbb{C}$ with the following properties:

- (i) They depend pointwise continuously on λ .
- (ii) For vanishing λ there is exactly one solution F^0 to $\mathfrak{h}^0(F) = 0$.
- (iii) The Jacobian $J_{ij}^{\lambda}(F) \equiv \frac{\partial \mathfrak{h}_{i}^{\lambda}}{\partial F_{j}}$ satisfies $\det[J^{0}(F^{0})] \neq 0$.

Then we have

- (a) For every ball B_r centered at F_0 there is a $\lambda_0 > 0$ such that for all smaller λ , \mathfrak{h}^{λ} also has a zero in B_r .
- (b) For every ball B_R centered at F_0 there is a $\lambda_0 > 0$ such that for all smaller λ , \mathfrak{h}^{λ} has at most one zero in B_R .

It can be checked easily, that these conditions are satisfied by all versions of the \mathfrak{h}^{λ} we defined above. The situation is depicted in Fig. 1. If the interaction is small, there is only one solution F_{phys} that is close to the non-interacting one. All others tend to infinity in the limit of small interaction, i.e. for $\lambda \to 0$ one can let the radius r tend to zero and R tend to infinity. This behavior suggests, that at least in some low coupling regime F_{phys} is indeed the physical solution, while all other fixed points are far away from the right result.

Proof. (a) First we note, that owing to the continuity with respect to λ given in (i) and the fact that \mathfrak{h}^{λ} is

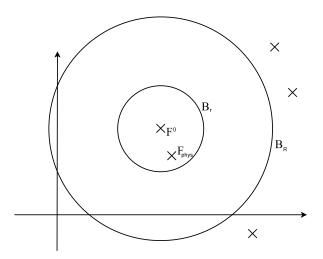


FIG. 1: Sketch of the position of the solutions of the interacting system relative to the non-interacting one F^0 . For small interaction there is one solution $F_{\text{phys}} \in B_r$ of Eq. (8), that is close to the non-interacting one F^0 . All others are outside B_R , while r tends to zero and R tends to infinity for the small coupling limit $\lambda \to 0$.

analytic then $J^{\lambda}(F)$ also depends continuously on both λ and F. So using (iii), we can restrict F and λ to be in a region around F_0 and 0, respectively, such that $J^{\lambda}(F)$ is close enough to $J^0(F^0)$ to have non vanishing determinant

$$\det[J^{\lambda}(F)] \neq 0.$$

We can assume, that the ball B_r named in (a) was already small enough to ensure this. Now we define

$$m \equiv \min_{F \in \partial B_r} \left\{ \|\mathfrak{h}^0(F)\| \right\} \neq 0,$$

where ∂B_r denotes the surface of B_r and $\|\cdot\|$ denotes the Euclidean vector norm. Now we are interested in the minimum of $\|\mathfrak{h}^{\lambda}\|$ on $\bar{B}_r = B_r \cup \partial B_r$. Clearly this minimum tends to zero for $\lambda \to 0$ since $\mathfrak{h}^0(F_0) = 0$. On the other hand, the minimum of $\|\mathfrak{h}^{\lambda}\|$ with respect to F on the surface ∂B_r tends to m > 0. So we can further restrict λ such that $\|\mathfrak{h}^{\lambda}\|$ has at least one local minimum in the interior of B_r . Let's call the location of one such minimum F_{phys} . Now that we know that this minimum is not on the surface of B_r , but in the interior, we can conclude

$$\frac{\partial \|\mathfrak{h}^{\lambda}\|^2}{\partial (\mathrm{Re}[F_i])}\bigg|_{F_{\mathrm{phys}}} = 0, \qquad \frac{\partial \|\mathfrak{h}^{\lambda}\|^2}{\partial (\mathrm{Im}[F_i])}\bigg|_{F_{\mathrm{phys}}} = 0.$$

This implies

$$\sum_{j} \frac{\partial \mathfrak{h}_{j}^{\lambda}}{\partial F_{i}} \mathfrak{h}_{j}^{\lambda *} \bigg|_{F_{\text{phys}}} = 0 \tag{9}$$

$$\sum_{j} J_{ji}^{\lambda}(F_{\text{phys}})[\mathfrak{h}_{j}^{\lambda}(F_{\text{phys}})]^{*} = 0.$$
 (10)

Now we have chosen λ and the size of the ball B_r such that the Jacobian has non-zero determinant. So the only way Eq. (10) can hold is if

$$\mathfrak{h}^{\lambda}(F_{\text{phys}}) = 0.$$

Proof. (b) We will do this proof in two steps. First we will chose an r > 0 such that we can show that a ball with this radius, centered at F^0 , contains no more than one solution (given λ is small enough). Then we will show, that for the same ball B_r and sufficiently small λ there is no solution in the set $B_R \setminus B_r$, where B_R is the ball named in (b).

Step 1. Define

$$M^{\lambda}[q_r] \equiv \int_0^1 J^{\lambda}(q_r(s)) ds,$$

with $q_r: [0,1] \to B_r$. Now $M^{\lambda}[q_r]$ can be decomposed into $J^0(F^0)$ plus a reminder, that vanishes in the limit of $\lambda \to 0$ and $r \to 0$. That allows us to fix r and restrict λ such that $\det(M^{\lambda}[q_r]) \neq 0$ for any map q_r . Now assume we have two zeros $F^1, F^2 \in B_r$ of \mathfrak{h}^{λ} . Then

$$\begin{split} 0 &= \mathfrak{h}^{\lambda}(F^2) - \mathfrak{h}^{\lambda}(F^1) \\ &= \int_0^1 \frac{d}{ds} \mathfrak{h}^{\lambda} [sF^2 + (1-s)F^1] \, ds \\ &= \sum_j (F^2 - F^1)_j M_{ij}^{\lambda} [q_r], \end{split}$$

where $q_r(s) = sF^2 + (1-s)F^1$. Now we use that $\det[M] \neq 0$ and therefore $F^2 = F^1$. So there is indeed no more than one solution in B_r .

Step 2. Due to the continuity with respect to λ granted in (i) the infimum of $\|\mathfrak{h}^{\lambda}\|$ with F restricted to $B_R \setminus B_r$ tends to the infimum of $\|\mathfrak{h}^0\|$ with the same restriction on F. This is not zero, since we assumed that $\mathfrak{h}^0(F) = 0$ has only one solution F^0 .

Convergence of iterative solutions

In practice, one solves Eq. (8) iteratively. That is, one starts with some initial guess F_0 , inserts it into the right hand side of $F = \mathfrak{g}(F)$ and obtains a new guess. Iterating this procedure defines a sequence

$$F_{i+1} = \mathfrak{g}(F_i). \tag{11}$$

A natural starting point for this is the non-interacting solution $F_0 = F^0$. The hope is that this sequence converges to a fixed point, i.e. a solution to the equations. A priori it is unknown if the calculation will actually converge, or if a fixed point obtained this way actually corresponds to the physical solution. So one may wonder, why GW and similar schemes do, in fact, work in many situations. We now show that, for weak coupling once again, a unique and convergent solution is guaranteed.

Theorem 2. If λ is not too large, iterating the equations according to (11), starting from the initial guess $F_0 = F^0$, converges to the physical solution $F_{\rm phys}$.

The connection to the physical solution is made through Theorem 1.

Proof. For $a,b,c\in\mathbb{R}$ we define the following transformation:

$$G'_0 = aG_0 W'_0 = bW_0 \Gamma'_0 = c\Gamma_0$$

$$G' = aG W' = bW \Gamma' = c\Gamma$$

$$\Sigma' = a^{-1}\Sigma \Pi' = b^{-1}\Pi \lambda' = c\lambda$$
(12)

with $a^2bc^2=1$. (We will use these transformations in this section only, to avoid confusion of the meaning of the primes with the derivatives as used earlier). This transformation leaves Eqs. (2), (3), (6) and e.g. (7) invariant:

$$F'_{i+1} = \mathfrak{g}'(F'_i).$$

We now apply the transformation with, say, $a = \lambda^{1/4}$, $b = \lambda^{1/2}$, $c = \lambda^{-1/2}$. This way \mathfrak{g}' depends on λ explicitly and implicitly through a, b and c. Observe, that all coefficients appearing in \mathfrak{g}' tend to zero as $\lambda \to 0$. The same is true for the transformed starting values $F_0' = F^{0'}$. In this situation Banach's fixed point theorem can be applied for the map \mathfrak{g}' defined in a vicinity of the non-interacting solution $F^{0'}$. We conclude, that for small λ the transformed quantities tend to a fixed point. Since the transformation can be inverted, this remains true for the original quantities. By Theorem 1 for small λ the solution F_{phys} is the solution nearest to the non-interacting one. Hence the fixed point obtained is indeed F_{phys} .

Generalizations of the Theorems

For the sake of clarity, we presented the Theorems 1 and 2 slightly less general than possible. However, these Theorems can be easily generalized. For example in the proof of Theorem 1 there was no reference to the fact that we are dealing with polynomials, rather only to the preposition that the functions are analytic in the region of interest (i.e. \bar{B}_r and \bar{B}_R). A consequence of this is that Theorem 1 also applies to vertex equations that are

more sophisticated than just polynomials. It also allows us to use Dyson's equations in the 'solved' form, e.g. $G = (1-G_0\Sigma)^{-1}G_0$, if we restrict Σ and Π (and hence B_R and B_r) such that the introduced poles are outside \bar{B}_R and \bar{B}_r . This restriction is natural in the non-interacting limit since then the self-energy and polarization should tend to zero. Similarly, one can extend the proof of Theorem 2 to these solved Dyson equations (note that they still obey the symmetry used in the proof). Also the restriction of choosing the non-interacting solution as starting point is actually not necessary, though it renders the proof much simpler; it can be shown that the fixed point becomes attractive in the small coupling limit, even without the restriction of starting from zero self-energy and polarization.

NUMERICAL INVESTIGATION

Numerical checks of the above theorems were performed for both self-consistent GW and the Starfish algorithm. As already mentioned, obtaining all solutions is in practice only possible for very small N. Hence for GWwe set N=2 and for Starfish we use N=1. Plotted in Fig. 2 are all the solutions for these algorithms, as a function of λ . The numerical input, in this case G_0 and W, were chosen to be random complex numbers, and Wwas kept fixed (which is common practice for real GWcalculations). As mentioned earlier, there are 6 solutions in the GW case. Of these, 5 tend to infinity and the remaining solution tends to G_0 as $\lambda \to 0$. This is a visualization of Theorem 1. For Starfish, 2 of the 3 solutions tend to a constant. This may seem to be in violation of the theorem, but in this case the vertex Γ (and therefore F) diverges.

We can also examine the domains of convergence for both of these algorithms. For N = 1 we fix $W_0 = 1$ and $\lambda = 1$ and plot the region of convergence of G_0 for the fully self-consistent GW and Starfish algorithm (for which W is also computed self-consistently) in Fig. 3. It can be observed that the region of stability shrinks for the higher-order method. Also noteworthy is that the region has a fractal boundary (this may be unsurprising since for case $G_0W = 1$ the domain is the Mandelbrot set). Perhaps more interesting is the region of starting points for which the algorithms converge. These are plotted in Fig. 4 for the same W_0 and λ but this time with $G_0 = 1 + i$ for GW and $G_0 = 1/4 + i/4$ for Starfish, and with a variable starting point for G. Once again the region of convergence is smaller for Starfish, but in both cases only one solution is found, irrespective of the starting point. This is a numerical confirmation of Theorem 2. Note that for GW a situation was picked, where the non-interacting starting point does not lead to convergence. Hence this can be considered a large coupling situation. But still there seems to be only one stable

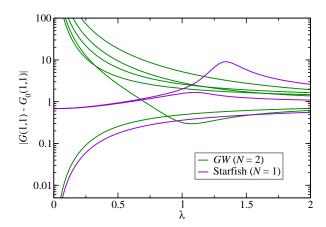


FIG. 2: Plot of the distance of a matrix element of the Green's function to the non-interacting one versus the coupling strength λ for all possible solutions of GW with N=2 and Starfish with N=1, for random G_0 and W. W is kept fixed in both cases. Always only one solution tends to the non-interacting one for the weak coupling limit.

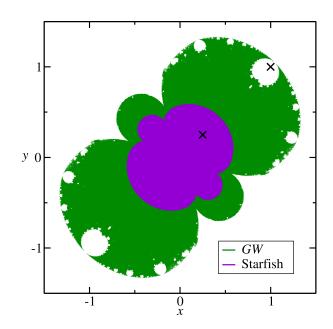


FIG. 3: Domain of convergence of GW and Starfish with N=1 for input values of G_0 in the complex plane, when using the non-interacting solution as starting point. Here $W_0=1$ and $\lambda=1$. The crosses mark the chosen G_0 for investigating the starting point dependence while fixing G_0 , see Fig. 4.

fixed point. The boundary of the region is also fractal (this corresponds to the Julia set).

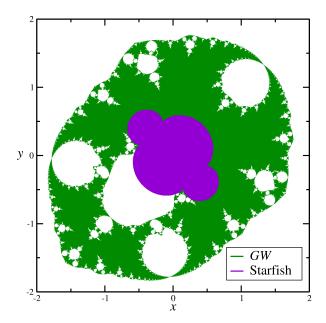


FIG. 4: Domain of convergence of GW and Starfish with N=1 for different starting points of the fixed point cycle. The values of G_0 are fixed to 1+i for GW and 1/4+i/4 for Starfish, as indicated by the crosses in Fig. 3.

CONCLUSIONS

We have argued that truncating Hedin's equations to some order yields systems of polynomial equations which have a very large number of solutions. As an example of this, the Starfish algorithm was introduced which includes vertex corrections beyond GW and consequently has even more fixed point solutions. The number of solutions tends to infinity as either the order of truncation or N tends to infinity, reflecting the inherent problem of solving Hedin's equations as a functional differential equation. Two theorems were presented that shed some

light on the general behavior of these fixed points. In particular we have shown, that there is exactly one solution that tends to the non-interacting case for small coupling, while all others are divergent in this limit. Numerical tests of self-consistent GW and the Starfish algorithm for small N demonstrated that the system also converges uniquely to one fixed point even for fairly large coupling. Furthermore, the region of stability may be fractal in nature, indicating that finding simple necessary and sufficient conditions for ensuring convergence of GW calculations $a\ priori$, may be impossible.

We thank Lucia Reining, Martin Stankovski and Ralph Tandetzky for useful discussions.

- [1] F. Aryasetiawan and O. Gunnarsson. arXiv:cond-mat/9712013v1, 1998.
- [2] Giovanna Lani, Pina Romaniello, and Lucia Reining. Approximations for many-body green's functions: insights from the fundamental equations. <u>New Journal of Physics</u>, 14(1):013056, 2012.
- [3] L. Hedin. Phys. Rev., 139:A796, 1965.
- [4] We stress that W and W_0 are two-point objects contrary to the interactions being four point objects.
- [5] Introducing two more integrations and using the bare vertex one can rewrite Eqs. (2), (3) such that they involve tensor contractions only. That way, all our equations of interest Eqs. (1)-(7) are basis-independent, though the bare vertex becomes more complicated in a different basis. Also the coupling constant λ no longer appears anywhere else than in the definition of the bare vertex Γ₀.
- [6] L. Hedin and S. Lundqvist. Effects of electron-electron and electron-phonon interactions on the one-electron states of solids. In Solid State Physics, Volume 23. Academic Press, New York, San Francisco, London, 1969.
- [7] W. Fulton. <u>Intersection Theory</u>, no. 2 in <u>Ergebnisse der Math</u>. Springer-Verlag, 1984.
- [8] B. Buchberger. Gröbner bases: An algorithmic method in polynomial ideal theory. In <u>Multidimensional Systems</u> Theory. New York: van Nostrand Reinhold, 1982.